Medical Knowledge Representation System

David BUCHTELA, Jan PELESKA, Miroslav ZVOLSKY, Jana ZVAROVA
EuroMISE Center and Center for Biomedical Informatics, Department of Medical Informatics, Institute of Computer Science AS CR v.v.i., Pod Vodarenskou vezi 2, 182 07 Prague 8, Czech Republic

Abstract. The aim of this article is to present a design of a Medical Knowledge Representation System (MEKRES). The system automatically offers relevant formalized knowledge by extended GLIF (Guidelines Interchange Format) models to participants (patient, physician, operator, …) on the basis of acquired data. This selection algorithm is based on key attributes and cooperation with knowledge ontologies.

Keywords. Knowledge Representation, GLIF model, Ontology

Introduction

Nowadays the conceptual knowledge (concepts, targets, actions, …) plays more prominent role. Conceptual knowledge is related to given task but expressed on higher level of abstraction and relatively independent of the form which is used by a knowledge system or expert. These factors are related to the creation and development of structured methodologies which support model based knowledge mining and using.

The aim of this article is to present a design of a Medical Knowledge Representation System which is based on GLIF models of branch knowledge. The GLIF models are extended by key attributes. The MEKRES offers the participants (patient, physician, operator, …) relevant model(s) based on acquired data.

1. GLIF model

The GLIF (Guideline Interchange Format) model is a result of collaboration among Columbia University, Harvard University, McGill University and Stanford University. The main goal of GLIF was to enable sharing of guidelines among institutions and across computer applications [3].

GLIF specifies a process-oriented model for guidelines representation and syntax for guidelines utilization in software systems as well as for their transport. GLIF guidelines are mostly given as a flowchart representing a temporarily ordered sequence of steps. The nodes of the graph are guideline steps and edges represent continuation...
from one step to the other one (see fig.1). Guideline steps are an action step, decision step, branch and synchronization steps and a patient state step.

- **Action steps** specify clinical actions that are to be performed. It can be an application of some therapy, carrying out some examination or measurement etc. Action step also may name sub-guidelines (subgraph), which provide a detail for the action.

- **Decision steps** are used for conditional branching. There are two kinds of decision steps: **Case step** is used, when branching is determined by evaluation of defined logical criteria based on data items. **Choice step** is used when the decision cannot be precisely specified in guidelines themselves and decision should be made by the user.

- **Branch** and **synchronization steps** enable concurrency in the model. Guideline steps that follow branch step can be performed concurrently. Branches with root in branch step eventually converge in a synchronization step. In this step all branches are synchronized after evaluation of synchronizing condition.

- **State step** characterizes a surveyed object state after execution of the previous step or in the beginning of the model.

**Figure 1.** Graphical symbols of GLIF model

1.1. **Criteria of conditions**

The decision step specifies several criteria of condition for each decision option:

- The **strict-in** criterion is used to specify a decision condition that could be computed automatically (for example if systolic blood pressure is 130 or greater). If a strict-in is true then the control flows to the guideline step that is specified by that decision option’s destination.

- The **strict-out** criterion is analogous to an absolute contraindication (for example if a patient is gouty he could not be cured by thiazides diuretics). If a strict-out is true then the decision option’s destination is forbidden.

- The **rule-in** criteria rank a choice as the best among several options. For example, when there are competing diagnoses for a disease, a pathognomonic
condition would be a rule-in for the disease. This criterion is analogous to conditions favouring the use (indications).

- A **rule-out** takes precedence over rule-in when ranking options. If an option contains both a rule-in criterion and a rule-out criterion, and both are evaluated as true, then that option should be the last choice. This criterion is analogous to contraindications.

The strict-out criterion is evaluated at first. If strict-out criterion is evaluated as true the rest of the criteria is not evaluated. This option is forbidden. In the opposite case the strict-in criterion is evaluated. If the strict-in criterion is false too, the rule-in and rule-out criteria are evaluated. The ranking of rule-ins and rule-outs is left to the users who may use their clinical judgment or may develop their own ranking schemes.

1.2. GLIF implementation

GLIF model is graphical so it is necessary to code it in a formal language. These formats are available for implementation of GLIF model and its parameters:

- **GELLO** (*Guideline Expression Language*) – the object oriented query and expression language for decision support [3].
- **RDF** (*Resource description framework*) – the built-in form of saved knowledge base in a Protége system [2]. The Protége is a development environment used for knowledge system design. The environment enables knowledge ontology construction, data inserting and user formatting of input and output forms.
- **XML** (*eXtensible Markup Language*) – The encoded model consists of a sequence of guideline steps. Some attributes of a guideline step contain next guideline steps. It enables sequential representation of a graph structure in the guideline language.

2. Knowledge ontology

Knowledge ontologies follow the development in an artificial intelligence in an area of knowledge representation. Ontologies are thoroughly taken in logic theories and their link to real objects (instances) is relatively opened. The ontology is primarily used to describe concepts (classes) and not to describe facts about concrete objects. Classes (concepts) and relations are systematically defined by a formal language [4].

The class is the base of the knowledge ontology and describes a set of objects. In some formalisms the class corresponds to a concept or a category and closely corresponds to a frame which is the base construction of any artificial intelligence systems. The ontology classes don’t contain procedural methods in opposite to classes in object-oriented models.

On a set of classes there is defined a hierarchy (taxonomy). The philosophical view of ontology sometimes requires a strict tree structure but all of the main ontology languages support multiple heredity.

The individuum corresponds to a concrete object of the real world and is the opposite of the class in a way. The individuum can be inserted into the ontology without a link to any class.
3. Medical Knowledge Representation System

3.1. Knowledge Representation Model

Knowledge Representation Model (KREM) of the whole system is based on branch knowledge formalisation through the use of the GLIF model. The formalisation process, i.e. construction of the graphic GLIF model of knowledge contained in free text and the model coding into the formal language (XML)[1], is illustrated in figure 2.

In the stage of a GLIF model construction from a free text it is important to find a logical and process structure of knowledge, all fundamental parameters and their interrelationships. The result of this stage is a graphic GLIF model corresponding to the knowledge in the text. The construction stage is the most important and difficult of all stages.

In the stage of GLIF model coding the graphic model of knowledge is coded into XML. Some steps of the encoded model contain description of next steps (next options). It enables sequential representation of a graph structure [1].

Table 1. Key attributes of knowledge model

<table>
<thead>
<tr>
<th>Key attribute</th>
<th>Attribute description</th>
</tr>
</thead>
<tbody>
<tr>
<td>branch</td>
<td>Branch described by the GLIF model – e.g. cardiology</td>
</tr>
<tr>
<td>BID</td>
<td>Branch identification – e.g. International Classification of Diseases (ICD)</td>
</tr>
<tr>
<td>user</td>
<td>User of system to whom the GLIF model is primarily determined (patient, physician, operator, …)</td>
</tr>
<tr>
<td>status</td>
<td>The GLIF model validity</td>
</tr>
<tr>
<td>key</td>
<td>List of keys described by the GLIF model – e.g. blood pressure, diabetes, …</td>
</tr>
<tr>
<td>key_name</td>
<td>Name of the key</td>
</tr>
<tr>
<td>key_weight</td>
<td>Weight of the key – the GLIF model description rate of the key</td>
</tr>
</tbody>
</table>
The resulting GLIF model is extended by key attributes. The list of the key attributes is in the table 1. These key attributes are used in selection algorithm and they are coded in XML along with the GLIF model.

3.2. Selection algorithm

The principle of the Medical Knowledge Representation system (MEKRES) and the algorithm of the relevant GLIF model selection is illustrated in figure 3.

The selection algorithm can be described subsequently:

- For every specific participant (user, patient, physician, operator, …) \( p \) and his attributes \( A_p \), a set of all branches (areas) \( \exp(B) \) which corresponds to participant state (attributes \( A_p \)) is determined. This function \( o \) is determined using the knowledge ontology:

  \[ o : A_p \longrightarrow \exp(B) \], where \( B \) is a finite set of all branches.

- For each branch \( B \in o(p) \) a subset \( K(p, B) \) of recognized attributes (keys) of the branch \( B \) and affected by attributes of the participant \( p \) is determined.

- Models \( G_p \) are chosen from a finite set of all GLIF models \( M \) thus:

  \[ G_p = \{ g \in M : g \in K(p, B) \cap \text{key_weight}(branch(g), P(g)) \neq 0 \} \],

  where each GLIF model contains attribute \( \text{branch} \) and a finite set \( P \) of keys (attribute \( \text{key_name} \) and their weights (attribute \( \text{key_weight} \)).

- For each model \( G_p \) a general aggregate operator \( r \) is defined:
\[ r : \mathbb{G}_p \to \mathbb{R}, \text{ where } \mathbb{R} \text{ is the set of real numbers.} \]

The operator \( r \) can be for example defined as \( R_g = \sum_{k \in \mathbb{G}_S} \text{weight}(k) \). The participant is then offered the GLIF model with the highest relevance value \( R = \max(R_g) \) or a list of models ordered by the value of \( R_g \).

4. Conclusion

Designed medical knowledge representation system is based on the GLIF model which is the universal method of modelling of mainly procedural-oriented knowledge. The system offers a relevant GLIF model or an ordered list of models from a list of available formal models on the basis of participant attributes. Additional information describing the concrete situation gives the participant possibility of better decision.

Acknowledgement

The work was partially supported by the project 1M06014 of the Ministry of Education of the Czech Republic and by the research institutional plan AV0Z10300504 of the Institute of Computer Science AS CR v.v.i.

References